LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-650084

On the Contribution of Longwave Radiation to Global Climate Model Biases in Arctic Lower Tropospheric Stability

N. P. Barton, S. A. Klein, J. S. Boyle

February 18, 2014

Journal of Climate

Disclaimer

Title:

On the Contribution of Longwave Radiation to Global Climate Model Biases in Arctic Lower Tropospheric Stability

Authors:

Neil P. Barton[1,2]

Stephen A. Klein[2]

James S. Boyle[2]

Affiliation:

[1]Science Applications International Corporation

[2]Lawrence Livermore National Laboratory

Program for Climate Model Diagnosis and Intercomparison

Abstract:

Previous research found that Global Climate Models (GCMs) usually simulate greater lower tropospheric stabilities compared to reanalysis data. In order to understand the origins of this bias, we examine hindcast simulations initialized with re-analysis data of 6 GCMs and find that 4 of the 6 models simulate within 5 days a positive bias in Arctic lower tropospheric stability during the Arctic polar night over sea ice and land regions. These biases in lower tropospheric stability are mainly due to cold biases in surface temperature, as very small potential temperature biases exist aloft.

In the hindcast runs, inter-model differences in downward longwave radiation at the surface explain a significant fraction of inter-model differences in polar night surface temperatures. Both clear sky and cloud radiative effects contribute to these longwave radiation differences. An important cloud property is the frequency of a cloud with liquid water path greater than 20 g m^{-2}. Inter-model spread in this quantity has a statistically significantly positive correlation to the inter-model spread of surface temperatures and longwave radiation.

These models were also analyzed in AMIP mode to determine if hindcast simulations are analogy to free-running simulations. Similar winter lower tropospheric stability biases occur in 4 of the 6 models with surface temperature biases relating to the winter lower tropospheric stability values.

1. Introduction

Temperature inversions in the lower troposphere are a common feature of the Arctic climate (Serreze et al. 1992; Liu et al. 2006; Zhang et al. 2011), and these surface based inversions are more frequent and more stable during the winter months compared to summer months (Tjernstrom; Graversen 2009; Zhang et al. 2011). Regional and Global Climate Models (GCMs) have difficulties representing Arctic inversions (Dethloff et al. 2001; Tjernstrom et al. 2008; Boe et al. 2009; Kay et al. 2011; Medeiros et al. 2011; Pithan et al. 2013). Boé et al. (2009) found that many of the models in the third Coupled Model Intercomparison Project (CMIP3) have an overly stable lower troposphere when compared to reanalysis data. In further evaluating the Lower Tropospheric Stability (LTS) in the CMIP3 models, Medeiros et al. (2011) determined that the Arctic LTS of 21 GCMs and reanalysis data are bimodal. There was a stable mode over the Arctic Ocean and adjacent continents, and an unstable mode over the North Atlantic Ocean. Medeiros et al. (2011) found that about half the climate models examined had an Arctic averaged overly stable LTS because of the bias in the partitioning of the two modes and the other half had an Arctic averaged overly stable LTS because of the bias in LTS in the stable mode. Pithan et al. (2013) examined CMIP5 (Taylor et al. 2012) models and found an overly stable lower troposphere occurred in many of these updated GCMs.

Representing the Arctic LTS in climate models is important because the mean LTS may be related to the amount of Arctic climate change in an enhanced CO_2 world (Boe et al. 2009; Bintanja et al. 2011; Bintanja et al. 2012). Boé et al. (2009) found a linear relationship between Global Climate Models' (GCMs') mean Arctic LTS and the Arctic longwave (LW) feedback in the CMIP3 archive, and suggested that GCMs have an unrealistic Arctic negative LW feedback because GCMs have a too stable Arctic lower troposphere. Bintanja et al. (2011) examined how the mean LTS state affects Arctic climate change in an enhanced CO_2 world in the GCM EC-Earth by altering the stable boundary layer mixing parameterization. As expected, lower mixing values resulted in a more stable mean Arctic LTS, while more mixing resulted in a less stable Arctic LTS in the GCM. When increasing CO_2 in EC-EARTH with these different values of boundary layer mixing, the run with a more stable LTS

in the mean state had a greater amount of Arctic surface warming and a greater amount of sea ice loss. Bintanja et al. (2011) theorized that the model with a more stable LTS allowed less Arctic low-level heat to be released into the free troposphere, and hence greater surface temperature change at the surface.

These studies motive further study into the biases of GCMs in representing Arctic LTS and their origins. In particular, this paper explores three questions:

Do GCMs have biases in Arctic LTS during all months/seasons? Boé et al. (2009) and Medeiros et al. (2011) found that Arctic LTS in GCMs is too stable during the winter months compared to reanalysis data, and yet Kay et al. (2011) found a similar biases in Arctic LTS in July examining National Center for Atmospheric Research's (NCAR's) Community Atmospheric Model version 4 (CAM4).

What atmospheric level or levels may be causing an Arctic bias in LTS? LTS is defined as the temperature or potential temperature difference between a level above the inversion and the surface. The GCM bias may be due a temperature bias above the inversion, at the surface, or of a combination of biases at both levels. Chapman and Walsh (2007) and Svensson and Karlsson (2011) found that many models in the CMIP3 archive underrepresent the amount of downwelling LW radiation during the winter months, and hence a surface temperature bias exists. Are the Arctic LTS biases solely due to surface temperature biases or do biases in the temperature above the inversion contribute?

What are possible causes of the Arctic LTS bias? As mentioned in Bintanja et al. (2011), the stability is controlled by the boundary layer mixing in GCMs, and one hypothesis in why Arctic LTS is bias in GCMs is that the boundary layer mixing is incorrect. With traditional representations of the mixing in the stable boundary layer, mixing tends to shut off as stability increases which may cause the surface to decouple from the troposphere leading to runaway cooling of the surface (Derbyshire 1999; Steeneveld et al. 2006) and overly stable conditions may occur. Often times models may arbitrarily increase the amount of stable mixing to avoid surface temperature biases under stable conditions (Sandu et al. 2013), but such tunings may not be present in all models. An alternate but not exclusive explanation of why Arctic LTS is incorrect in GCMs, which we investigate in this paper, is that

there is a misrepresentation in the radiative driving of the surface temperature, particularly a misrepresentation of surface downwelling radiation, which may be impacted by clouds.

Previous case studies have demonstrated that clouds greatly regulate surface temperature in the Arctic (Shupe; Intrieri 2004), and models routinely have a difficult time representing Arctic clouds correctly (Tjernstrom et al. 2008; Klein et al. 2009; Morrison et al. 2009; Pithan et al. 2013). The increase in downward longwave radiation that occurs with clouds may be enhanced by increased longwave emission from water vapor which tends to be higher when clouds occur (Francis et al. 2005). If errors in clouds contribute to significant errors in the downward longwave radiation, then we need to determine which characteristics of clouds are most responsible. In fact, Pithan et al. (2013) suggested that a miss-representation of mixed-phase clouds during the Arctic winter relates to errors in the CMIP5 models' LTS.

In this study, we use GCM hindcast runs (Phillips et al. 2004) to analyze Arctic LTS in a subset of CMIP5 models. Because hindcasts are initialized with operational weather analyses, the analysis of hindcast runs allows for a determination of initial error in climate models. Multiple feedbacks may lead a GCM bias to worsen, but these feedbacks may also lead to compensating errors in which biases are reduced; either way hindcast runs aid in isolating the factors contributing to GCM biases. In addition, hindcast runs allows for a comparison with observations at specific time steps and locations, something that is not as easily done with the typical GCM diagnosis, which relies on comparisons of the statistics from decades of model output with observations or re-analysis. Furthermore, we analyze high frequency output from the hindcast simulations, which allows for a more process driven approach in determining biases. This is particularly helpful in this study since the Arctic surface is dominated by a bimodal distribution of net surface LW radiation apparent in hourly data (Stramler et al. 2011; Morrison et al. 2012). This bimodal distribution is masked by monthly averages, but may be important for simulating the Arctic climate.

In this paper, we first describe the hindcast modeling runs (Section 2.1). We also examine the same GCMs in a free running mode to relate to the hindcast modeling runs, and a description of these models is found in Section 2.2. Next, a description of operational analysis, reanalysis and in-situ data is presented (Section3), and we discuss methods and variables definitions (Section 4). We first present results of the hindcasts runs for areas poleward of 70°N (Section 5.1) and at Barrow, Alaska (Section 5.2). Then we present the results of free-running climate integrations of the same models (Section 5.3). Last, we discuss and reiterate the main conclusions of this paper (Section 6).

2. Model Runs

2.1. *Hindcast Runs*

Hindcast runs from the Transpose AMIP II (T-AMIP) experiment and internal runs at Lawrence Livermore National Laboratory (LLNL) are used in this paper. Transpose AMIP is a model intercomparison experiment for GCMs run in hindcast mode (Williams et al. 2013) with a goal to understand how biases grow in climate models from a well-initialized state. The hindcast method generally emphasizes model errors in the moist processes because the analyses used to initialize models constrain the large-scale dynamics to be close to observed analysis (Phillips et al. 2004; Xie et al. 2012).

The models we use in the Transpose AMIP experiment include: HadGEM2-A, IPSL-CM5A-LR, CNRM-CM5, and MIROC5 (Table 1). For each season, 16 hindcast runs were performed. The seasons started on the 15th of October 2008, January 2009, April 2009, and July 2009 at 00Z. The 16 hindcast runs for each season were started 30 hours apart from each other, produced 5-day forecasts, and 3 hourly output were archived. These runs were initialized by the European Center for Medium Range Forecasting Year of Tropical Convection (ECMWF-Y) analysis data. The model output from Transpose AMIP is available on the Earth System Grid federation (ESGf). More detailed information on the T-AMIP runs is presented in Williams et al. (2013).

In addition, the Community Atmospheric Model version 4 and 5 (CAM4 and CAM5) from NCAR were run at LLNL in hindcast mode. These runs were initialized

at 00Z every day from May 2008 to March 2010 and produced 5 day forecasts. These are the same runs used in Barton et al. (2012), Xie et al. (2012), and Ma et al. (2013). The initialization data for CAM4 and CAM5 are the same as the Transpose AMIP initialization data (i.e., the ECMWF-Y analysis). To be consistent in the data analysis, CAM4 and CAM5 runs are only analyzed during the seasons laid out by the Transpose AMIP runs. There are more data points in the CAM4 and CAM5 output because these models were initialized everyday instead of once every 30 hours. Official T-AMIP CAM4 model runs will be released on the ESGf in the future, but these runs were not available at the time of this study.

2.2. *AMIP Runs*

To compare with the hindcast results, we analyze free running or Atmospheric Model Intercomparison Project (AMIP) runs for the models listed in Table 1. AMIP runs have monthly-observed sea surface temperatures and sea ice concentrations as boundary conditions. We analyze daily output from the AMIP runs in which atmospheric composition emulates the 20th century (i.e., 20th century historical AMIP runs), and the output from these runs is available at ESGf under the CMIP5 (Taylor et al. 2012) archive.

3. Operational Analysis, Reanalysis, and In-Situ Data

To evaluate the hindcasts at a large-scale, we compare models to the ECMWF operational analyses for the Year of Tropical Convection (Waliser et al. 2012) (ECMWF-Y). These are the same data used to initialize the hindcast runs. Hindcast modeling runs were also examined against in-situ observations. We use the Climate Modeling Best Estimate (CMBE) data (Xie et al. 2010) at Barrow (e.g., the North Slope of Alaska (NSA)) for this analysis. The data from Barrow were collected from a long-term measurement site of the US Department of Energy's Atmospheric Radiation Measurement (ARM) program. Lastly, to evaluate the AMIP simulations at larger temporal averages than the ~2 years of ECMWF-Y, we use the ECMWF ReAnalysis Interim (ERA-I) output (Dee et al. 2011).

Because we evaluate the GCMs against the ECMWF-Y analysis, the question arises on how well the ECMWF-Y analysis represents the components of LTS. Table 2 displays the bias, Mean Absolute Bias (MAE), and Root Mean Square Error (RMSE)

between ECMWF-Y and the NSA LTS (Eq. 1), surface temperature, and potential temperature at 850 hPa. The CMBE observations use the available radiosonde data collected by the ARM program at its NSA site. This comparison was performed using hourly mean data every 3 hours during periods of T-AMIP runs and at the ECMWF-Y grid point which is closest to the NSA. The results are presented for all, polar night, and polar day periods (defined in Section 4.1).

The absolute bias for all measures is 1 K or less, with the largest bias at -1 K for LTS during the polar day. The largest MAE and RMSE occur for LTS during the polar night periods with values of 2.63 K (MAE) and 3.55 K (RMSE). The LTS error is a combination of errors at the surface and the potential temperature at 850 hPa. MAE and RMSE errors are always larger at the surface compared to errors at 850 hPa. Statistically, the ECMWF-Y compares remarkably well to the NSA data though a more thorough comparison of the ECMWF-Y analysis with different in-situ at different locations would be beneficial. Of course, small errors in the ECMWF-Y may arise because data from the operational weather sounding at NSA are used by the ECMWF data assimilation package. With these known caveats, we still feel that it is beneficial to analyze the hindcast simulations with ECMWF-Y operational analysis over the Arctic domain.

4. Methods

4.1. *Temporal Period of Analysis*

Because the controls on Arctic temperature structure differ depending on whether or not solar energy is available, we choose to analyze the hindcast runs not based on the season as defined by the T-AMIP runs, but based on if the sun was above or below the horizon. For each time step, data averaged poleward of 70°N of the hindcast simulation is labeled either as Polar Night or Polar Day based upon an average Solar Zenith Angle (SZA) calculated for all modeled longitudes at a latitude of 80°N. For the comparison with the CMBE-NSA data, the SZA, hence Polar Night and Polar Day, is calculated at each time step using the latitude and longitude at Barrow. The AMIP runs are analyzed during the same temporal periods that overlap with the ERA-I product (January 1979 to December 2008), and Polar Night/Day is

224 defined the same way as the T-AMIP runs, but using daily output instead of 3 hourly
225 output.

226 **4.2. *Lower Tropospheric Stability (LTS)***

227 LTS is defined as the temperature or potential temperature difference at a
228 height above the boundary layer (x) and at a layer near the surface (y) (i.e.,
229 $LTS = \theta_x - \theta_y$). The surface is usually defined at 1,000 hPa for the lower latitude
230 oceans (Klein; Hartmann 1993). In the Arctic, the surface temperature is a more
231 ideal indicator of LTS than 1,000 hPa because the surface temperature can be lower
232 than the temperature at 1,000 hPa. In this study, we define the surface as the 2-
233 meter air temperature, which is a standard output for GCMs, operational analysis,
234 reanalysis products, and in-situ observations. We have verified that the main
235 conclusions would not change if surface skin temperature was used in place of 2-
236 meter air temperature.

237 In the mid-latitudes, the height above the boundary layer used for LTS
238 calculations is usually 700 hPa (Klein; Hartmann 1993). In the Arctic, the boundary
239 layer is not as deep as in the mid-latitudes and multiple heights have been used to
240 estimate LTS: e.g., 700 hPa (Barton et al. 2012), 850 hPa (Boe et al. 2009; Medeiros
241 et al. 2011), and 925 hPa (Kay; Gettelman 2009; Kay et al. 2011; Pavelsky et al.
242 2011). Vertical temperature profiles of the Transpose-AMIP runs and ECMWF-Y
243 analysis (not shown) show that the average height of the inversion is below 850 hPa
244 and 925 hPa; consequently, we define LTS as Eq. 1.

$$LTS \equiv \theta_{850} - \theta_{2m} \qquad 1$$

245 **4.3. *Longwave Cloud Radiative Effect***

246 In order to study the impact of clouds on the surface radiation budget during
247 the polar night and its relationship to LTS, we examine the models' longwave Cloud
248 Radiative Effect (CRE_{LW}). The Cloud Radiative Effect (CRE), or cloud radiative
249 forcing, is a measure of the clouds effect on the environment (Ramanathan et al.
250 1989). In this study we analyze the CRE_{LW} at the surface, defined as Eq. 2,

$$CRE_{LW} \equiv LW\downarrow(A) - LW\downarrow(0) \qquad 2$$

, where LW↓(A) is the surface downwelling longwave radiation for all sky conditions, and LW↓(0) is the surface downwelling longwave radiation assuming clear sky conditions.

While LW↓(A) and LW↓(0) are both readily available from standard model output, LW↓(0) is not available from the observations. Thus, we calculate it using the Rapid Radiative Transfer Model for GCMs (RRTMG) (Mlawer et al. 1997) for periods that temperature and moisture profiles were available from the radiosondes, which were launched every 12 hours at Barrow. In these RRTMG calculations, cloud and aerosol properties were set to zero, ozone concentrations were taken from the monthly ERA-I product, and CO_2 concentrations were taken from NCAR's Community Earth System Model CESM historical runs.

Defining CRE_{LW} with only downward (instead of net) fluxes follows previous research (Barton; Veron 2013; Shupe; Intrieri 2004), and is a measure of the instantaneous effect of clouds on the surface energy budget. CRE_{LW} values are generally positive because the LW↓ is greater with a cloud than without, and this represents a warming effect of clouds. During the polar night, we emphasize that CRE_{LW} is the total cloud radiative effect because of the lack of insolation.

4.4. *Cloud Properties*

If CRE_{LW} significantly relates to Arctic LTS biases, it is important to understand which cloud properties contribute to the relationship. Cloud properties that affect averaged CRE_{LW} include cloud cover amount (both frequency and amount when present), and cloud Liquid Water Path (LWP) and cloud Ice Water Path (IWP). Due to the lack of available model data, the effects of cloud particle size on longwave radiation (Garrett; Zhao 2006) are not considered in the present study. In the Arctic, the bimodal temporal distribution of hourly surface LW radiation relates to differences in humidity, and may relate to differences in clouds (Morrison et al. 2012; Stramler et al. 2011). The temporal distribution (i.e., frequency) of cloud properties are quantified for a comparison with surface temperature.

In addition, we compute the fraction of grid-boxes that LWP and IWP in the 3-hourly (or daily) output exceed specified thresholds, to roughly represent the fraction of time that the atmosphere contains an opaque cloud. The LWP and IWP

thresholds are determined by the LWP and IWP amount that affects surface LW↓, though we acknowledge that other cloud properties, such as vertical placement and temperature, also affect surface LW↓. Arctic clouds saturate in the longwave with LWPs near 30 g m^{-2} (Shupe; Intrieri 2004; Chen et al. 2006). An increase in downwelling longwave radiation at the surface is not expected when a cloud LWP increases beyond the LWP saturation value when the temperature and moisture conditions are the same. Even though clouds saturate in the LW with LWPs near 30 g m^{-2}, we choose a threshold of 20 g m^{-2} in this paper. This threshold is chosen because clouds with LWPs at 20 g m^{-2} (i.e. non-opaque clouds with still appreciable cloud emissivity) affect the surface downwelling longwave radiation, and hence surface temperature. Clouds with lower LWPs also affect the surface radiative budget, but the uncertainty of the LWP measurements is near 25 g m^{-2} (Westwater et al. 2001). We feel that the uncertainty of the measurement compared to the signal is too large if we defined an opaque liquid cloud with a lower threshold than 20 g m^{-2}, but our main results are not altered if the LWP threshold ranges from 10 to 30 g m^{-2}.

The IWP threshold for opaqueness is suggested by Ebert and Curry (1992). They modeled cloud emissivity as a function of cloud IWP and ice effective radius, and through their analysis, we choose an IWP threshold for 25 g m^{-2} of cloud opaqueness caused by ice. In Ebert and Curry (1992), clouds with IWPs at 25 g m^{-2} had emissivities ranging from 0.5 to >0.9 with effective radii ranging from 20 to 90 μm. The T-AMIP model output does not have ice cloud effective radii data to compare with the modeling results of Ebert and Curry (1992), but changing the IWP threshold to 40 g m^{-2} does not change the results of this paper. In addition, CMBE-NSA data do not have IWP observations so measurement uncertainty is not discussed. In addition, to separately computing the frequency that LWP and IWP individually exceed a threshold, we also analyze the frequency that either the LWPs are above 20 g m^{-2} or IWPs are greater than 25 g m^{-2} to determine a full opaqueness of the atmosphere due to clouds. In passing, we note we are using the 3-hourly grid-box mean LWP or IWP instead of computing whether an in-cloud value of LWP or IWP exceeds the threshold and then multiplying by the cloud fraction for the 3

hourly period. The difference in frequency of opaque cloud calculated with this alternate approach is likely very small.

5. Results

5.1. *Hindcast Biases Poleward of 70°N*

5.1.1. *Biases in LTS, Potential Temperature at 850hPa, and Near Surface Temperature*

For each 3-hour temporal period, variables were averaged poleward of 70°N, and then averaged for polar night and polar day periods. Averaged polar night and polar day biases of LTS, potential temperature at 850 hPa, and surface temperature are shown in Figure 1. Biases are more prominent during the polar night compared to the polar day for all models except CNRM-CM5, which has small biases in both the polar night and polar day. During the polar night, CAM4, CAM5, HadGEM2-A, and IPSL-CM5A-LR all have a more stable Arctic lower troposphere compared to the ECMWF-Y analysis, in agreement with the previous findings that climate models tend to have overly stable conditions in Polar regions. In contrast, the MIROC5 model exhibits very different behavior by drifting towards a less stable lower troposphere and has on forecast day 5 a LTS bias of -2.7 K during the polar day and -6.2 K during the polar night. In most models, the bias nearly saturates by the hindcast day two and the biases are not much larger in during days three to five. Two exceptions are CAM5 and MIROC5, in which biases continue to grow with forecast day. Relatively small biases in the day one hindcast in CAM4, CAM5, HadGEM2-A, CNRM-CM5 suggest that the initialization is successful setting LTS to the analysis value.

The biases θ_{850} are very small compared to the biases in the LTS during the polar day and polar night periods, and the biases in surface temperature are of a similar magnitude as the LTS biases for all models in both the polar day and polar night. The relative smallness in the hindcasts of biases in θ_{850} is perhaps not surprising given that the processes governing its evolution evolve on a longer time and space scales than those associated with surface temperature. Indeed, analysis of the AMIP integrations shows that at climate timescales biases in θ_{850} can contribute to biases in LTS in some seasons in some models (see Figure 10 below). During the

polar night, CAM4, CAM5, HadGEM2-A, and IPSL-CM5A-CM5-LR all have surface temperatures lower than the ECMWF-Y analysis, which relates to the LTS value to be larger than ECMWF-Y. MIROC5 has a warmer surface temperature than the ECMWF-Y analysis, and the LTS values are lower than the ECMWF-Y analysis. All models have larger biases during the polar night compared to the polar day, and MIROC5 has the largest biases compared to the other models during the polar day.

Medeiros et al. (2011) stated that Arctic LTS is a bimodal distribution with higher LTS values occurring over the Arctic sea ice and land, and lower LTS values occurring over the Arctic open ocean. In addition, Pavelsky et al. (2011) found that mean annual sea ice concentrations correlate very well with December, January, February mean Arctic LTS ($r = 0.88$). Because the mean Arctic LTS state is dependent on sea ice, are the biases also dependent on surface type? In Figure 2, the Arctic LTS, potential temperature at 850 hPa, and near surface temperature biases are displayed averaged over regions of sea ice, land, and open water for regions poleward of 70°N during the polar night period. The polar night period is only shown because of the relatively large biases compared to the polar day. The largest biases occurred over sea ice covered regions compared to regions of land and open water. The relatively small biases over open water are expected because these models are forced with observed SSTs.

Even though the LTS biases over sea ice are larger than over land, these biases are generally of the same sign in a given model. This suggests that the biases are not caused by processes in the sea ice or land. The IPSL-CM5A--LR model is an exception in which over sea ice, the LTS drift is toward a more stable lower troposphere and there is little LTS bias over land. As displayed in Figure 1, the LTS biases over sea ice is very much related to the near surface temperature bias. Hereafter we focus on polar night over sea ice when examining the hindcast modeling runs because of the relative larger biases.

5.1.2. *Model Spread in Near Surface Temperatures*

5.1.2.1. *Comparison with LW Radiation*

Polar night biases in LTS very much relate to biases in the near surface temperature. The correlation coefficient between these area-averaged variables for

each model and hindcast day is -0.993. Because of the connection between Arctic LTS and surface temperature bias and the relatively small biases at 850 hPa, we focus on understanding the spread in surface temperatures across the models and hindcast days.

The spread in surface temperature is very much related to the spread in surface downwelling longwave radiation (LW↓) during the polar night (Figure 3). Without solar radiation and relatively inefficient other sources of energy, this reflects the dominance of longwave radiation on the surface energy balance over sea ice in the polar night. It suggests that the surface temperature in these hindcast runs is coming into radiation balance with the longwave radiation incident upon it from the atmosphere. Indeed, the slope of the line is 3.8 W m^{-2} K^{-1} which is approximately equal to slope of the dependence of upward LW emission using the typical surface temperatures from the hindcast models (i.e., $\frac{\partial\, \sigma T_{surface}^4}{\partial\, T} = 4\sigma T_{surface}^3$, where σ is the Stefan Boltzmann constant and $T_{surface}$ is the surface temperature).

Is the inter-model spread in LW↓ associated with spread in clear-sky or cloudy-sky conditions? Figure 3, which examines the intermodel spread in clear sky and CRE_{LW}, shows that there are statistically significant (defined by p-values less than 0.05) relationships of these two variables with near surface temperature. Models with higher surface temperatures, tend to have larger clear sky LW↓ and their clouds are more radiatively opaque. The spread in LW↓ in the models is ~45 W m^{-2} and the spread in clear sky LW↓ and CRE_{LW} is ~25 W m^{-2}. An interesting observation of the CRE_{LW} scatter plot is that for each model, except IPSL-CM5A-LR, the day 1 CRE_{LW} forecast is more different from forecast in the other days, compared the differences in the day 2, day 3, day 4, and day 5 forecast with themselves. This suggests that the cloud characteristics for each model have not completely spun up in the day 1 forecast.

Because these models were initialized with the same temperature and moisture profiles one would expect the spread in clear-sky LW↓ to be relatively small, at least initially. Thus, the relatively large spread in clear sky LW↓ is worth further investigation, and we examine the hour 3 to 21 hindcast output. The hour 3

405 spread in clear sky LW↓ is 11 W m^{-2}, which is only about 4 W m^{-2} less than the
406 average day 1 spread, and 40-50% of the inter-model spread at day 3. Why is there
407 an 11 W m^{-2} difference in clear sky LW↓ at the first time that output was archived?
408 This spread may be due to differences in temperature and humidity that develop
409 during the first 3 hours of the hindcast, radiation code, and/or other factors that
410 affect LW radiation, such as aerosols, trace gasses, or even the vertical resolution of
411 the model. To test how much of the initial difference is due to differences in the
412 temperature and humidity profiles present even in the first day, we ran RRTMG
413 using temperature and humidity profiles averaged poleward of 70°N for each model
414 and compared clear sky LW↓ calculated by RRTMG to that computed by model itself
415 as reported in the model output (Figure 4). The RRTMG calculations set clouds to
416 zero, and ozone and other gases were the same for each model, and the vertical
417 grids of each model were first interpolated to the ECMWF-Y grid. For reference, we
418 also include the result of RRTMG calculations using the temperature and water
419 vapor fields from ECMWF-Y analysis in Figure 4. At forecast hour 3, these RRTMG
420 calculations had a clear sky LW↓ spread of 4 W m^{-2}, less than half of the 11 W m^{-2}
421 spread in the model-calculated LW↓. This suggests that the initial spread in clear sky
422 LW↓ is mostly not due to differences in temperature and moisture profiles and that
423 the inter-model spread of modeled LW↓ in the first day is not being driven by initial
424 temperature and moisture differences. From this evidence as well as the fact that
425 inter-model spread in LW↓ from early in the hindcasts is correlated with inter-
426 model spread in surface temperatures, we suggest differences in the way a climate
427 model determines clear-sky LW↓ (resulting from differences in either the other
428 radiation code input such as trace gasses, the vertical resolution of the radiation
429 code, or the radiation code itself) are an important contributor to inter-model
430 spread in surface temperatures in the hindcasts. This motivates further research
431 into the way models treat longwave radiation in the Arctic perhaps through
432 structured radiation code intercomparisons (Oreopoulos et al. 2012).
433 In passing, we note that most of the inter-model spread of 4 W m^{-2} in clear
434 sky LW↓ calculated with RRTMG at hour 3 results from the MIROC5 model.
435 Specifically, the MIROC5 has at ~165 W m^{-2}, whereas the values in other 5 models

436 are in the range of 160 to 163 W m^{-2}. This points to significantly larger drifts in
437 MIROC5 at hour 3 in the temperature and moisture profiles than in the other
438 models, which might indicate problems with model initialization in MIROC5.
439 **5.1.2.2.** ***Comparison with Cloud Properties***
440 Figure 3 also shows that there is a statistically significant relationship
441 between CRE_{LW} and surface temperature, which motivates an analysis of which
442 cloud properties contribute to the inter-model spread of near surface temperature.
443 In Figure 5, we analyze the spread in total cloud cover, LWP, IWP, frequency of
444 clouds with LWPs greater than 20 g m^{-2}, frequency of clouds with IWPs greater than
445 25 g m^{-2}, and the frequency of clouds with LWPs greater than 20 g m^{-2} or IWPs
446 greater than 25 g m^{-2} in relation to the models surface temperature. All data are
447 averaged over sea-ice regions during polar night. The linear relationships between
448 surface temperature and total cloud cover or IWP are not statistically significant at
449 the 0.05 level. CAM5 has relatively high values of total cloud cover, but low values of
450 surface temperature; while MIROC5 is near the median amount of total cloud cover,
451 but has high surface temperatures. The IPSL-CM5A-LR, HadGEM2-A, and MIROC5
452 models have relatively high values of IWPs, but the range in surface temperatures
453 among the models is close to 10 K.
454 The linear trend in the averaged LWP and the surface temperature is
455 statistically significant with an r value of 0.579 and a p value less than 0.05. Larger
456 LWPs occur in the models with higher surface temperatures. One exception is CAM4,
457 which has higher mean LWPs compared to the other models, but lower surface
458 temperatures. There are two models (CAM5 and HadGEM2-A) that have averaged
459 LWPs close to zero in the polar night. Arctic observations show liquid in clouds
460 occur throughout the polar night over sea ice (Shupe et al. 2011). The models' LWPs
461 are not correlated with the models' IWPs. For example, IPSL-CM5A-LR and
462 HadGEM2-A have high IWP values, but low LWP values during this period of
463 analysis. In addition, CAM4 has higher values of LWP compared to other models, and
464 a median value of IWP.
465 For CAM4, we note that there is a disconnect between mean LWPs, LW↓, and
466 CRE_{LW} between downwelling longwave radiation. For example, CAM4 CRE_{LW} values

467 are the second lowest compared to the other models, but the LWP values are the
468 highest. One possible explanation is that the polar night averages are masking high
469 frequency variability. If very large LWPs occur at only a few time steps, the mean
470 LWP would be relatively high, but the CRE_{LW} would be low. The temporal
471 distribution of LWPs is more related to mean CRE_{LW} than the mean LWP, and
472 motivates examination of the measures related to the frequency of opaque clouds
473 The regression between surface temperature and the frequency of clouds
474 with LWPs greater than 20 g m^{-2} is statistically significant, and the r value of 0.776 is
475 close to the r value between surface temperature and CRE_{LW} (r = 0.821). This
476 suggests that the surface radiative effect over sea-ice of clouds in these hindcast
477 simulations is largely controlled the frequency that a liquid cloud occurs during the
478 polar night. In these model runs, the frequency that a cloud occurs with LWPs
479 greater than 20 g m^{-2} relates to the surface temperature to a greater extent than to
480 the mean cloud LWP during the polar night. CAM4 is still an outlier in the analysis
481 with a relatively high frequency of clouds with LWPs greater than 20 g m^{-2} but low
482 surface temperatures. CAM5, HadGEM2-A, and IPSL-CM5A-LR have a frequency of
483 clouds with LWPs greater than 20 g m^{-2} that is less than 20%. The CNRM-CM5 model
484 has a frequency near 40 to 50% and surface temperatures near the ECMWF-Y
485 analysis. The MIROC5 model has frequencies slightly above 50%, which relates to a
486 high mean surface temperature compared the other models and ECMWF-Y analysis.
487 The regression between surface temperature and the frequency of clouds
488 with IWPs greater than 25 g m^{-2} is not statistically significant, and as with the mean
489 IWP comparison, the models with a high frequency of clouds with IWP greater than
490 25 g m^{-2} are not the same models with a high frequency of clouds with LWPs greater
491 than 20 g m^{-2}. A lack of relationship of CRE_{LW} with the frequency of IWP greater than
492 25 g m^{-2} may reflect the fact high-altitude ice clouds may less efficiently change the
493 surface LW↓ than liquid or mixed-phase clouds which tend to be closer to the
494 surface. The lack of a relationship with IWP means that the reason that the
495 frequency of clouds with LWPs greater than 20 g m^{-2} or IWPs greater than 25 g m^{-2}
496 significantly relates to surface temperature (Figure 5f) is due to the relationship
497 with the frequency of LWPs greater than 20 g m^{-2}.

5.2. *Hindcast Biases at Barrow*

5.2.1. *Biases in LTS, Potential Temperature at 850hPa, and Near Surface Temperature*

A benefit of hindcast climate model simulations is that these runs can be compared to in-situ observations at specific times. This has been done in previous Arctic research (Xie et al. 2008; Kay et al. 2011; Liu et al. 2011; Barton et al. 2012), in which biases have been observed and possible solutions to the problems put forth. In addition, we did not examine LW↓ and cloud LWPs in the ECMWF-Y analysis because these variables are parameterized, and in-situ data allows for a comparison of these variables that are not a result of physical model parameterizations.

Variables for each model are examined at the grid point closest to the North Slope of Alaska (NSA) site at Barrow and compared to the Climate Model Best Estimate (CMBE) data. Compared to the analysis conducted poleward of 70°N with the ECMWF-Y analysis, the biases for the climate models are not as large, but comparable to the polar night land biases for 5 of the 6 models with all models having the same sign except HadGEM2-A (Figure 6). Similar biases are not necessarily expected at the NSA because the analysis is at local region compared to averages poleward of 70°N, the CMBE-NSA data are located over land, and differences in model horizontal resolution.

The analysis at the NSA shows the similar linear relationship between LTS and surface temperature in the polar night (Figure 6). The linear regression of the models' hindcast value for each day between LTS and near surface temperature is -0.979. Similar to the analysis poleward of 70°N, the MIROC5 model has higher surface temperatures and a less stable lower troposphere compared to CMBE-NSA data. CAM4, CAM5, and IPSL-CM5A-LR have LTS values slightly larger than the mean LTS value from the CMBE-NSA data similar to the analysis poleward of 70°N, but many of these models' hindcast days lie in the 95% confidence interval of the CMBE-NSA data. The HadGEM2-A model has the opposite sign in the CMBE-NSA results compared to the analysis performed poleward of 70°N (i.e., Figure 2). The linear relationship between LTS and surface temperature leads to the conclusion that an

analysis of the relationship between the surface temperature biases and other variables at NSA would be relevant to the analysis for the land and sea ice covered Arctic.

5.2.2. *Model Spread in Near Surface Temperatures*

5.2.2.1. *Comparison with LW Radiation*

At the NSA, there are observed values of LW↓. Clear sky LW↓ and CRE_{LW} are calculated with RRTMG, as explained in Section 4.3. The linear relationships between LW↓, clear sky LW↓, and CRE_{LW} with surface temperature is examined at the NSA (Figure 7). Similar to the analysis poleward of 70°N over sea-ice, the surface temperature in the models is very much related to the downwelling surface radiation at the NSA during the polar night. The slope of this line is 3.8 W m^{-2} K^{-1}, which is again very close to theoretical value at the models' temperature range. This suggests that when the sun is below the horizon in the Arctic, surface temperature drifts toward (longwave) radiative equilibrium in the hindcast runs. The linear relationships between clear sky LW↓ and CRE_{LW} with surface temperature are statistically significant with r values of 0.882 and 0.701 representatively. Clear sky LW↓ and CRE_{LW} are related in these models, similar to the analysis performed poleward of 70°N, with the IPSL-CM5A-LR model being an exception.

5.2.2.2. *Comparison with Cloud Properties*

We further analyze which cloud properties relate to the spread in surface temperatures in the hindcast models (Figure 8). Similar to the analysis poleward of 70°N, total cloud cover and mean IWPs do not significantly relate to surface temperature at the NSA. CAM5 has the highest cloud cover amount, but the surface temperatures are relatively low. MIROC5 and IPSL-CM5A-LR have similar cloud cover amounts, but surface temperatures differ near 8 K.

Unlike the analysis poleward of 70°N, the linear relationship between near surface temperature and mean LWP during the polar night is not statistically significant (Figure 7, r = 0.085, p = 0.654). Again, CAM4 is an outlier in the analysis with relatively large LWPs but relatively low surface temperatures compared to the other models. MIROC5 has the next largest values of cloud LWP (compared to CAM4), and MIROC5 has the largest values of surface temperatures. CAM5 and IPSL-

CM5A-LR both have mean LWP values lower than the CMBE-NSA data and lower surface temperatures. Many CNRM-CM5 values of LWP and surface temperature lie in the uncertainty range of the CMBE-NSA data. The HadGEM2-A model has relatively low values of LWP, but relatively large surface temperatures. In addition, the HadGEM2-A model has relatively high values of IWP compared to the other models, which may be aiding in the relatively high temperature values.

When examining the frequency of clouds with a LWP threshold of 20 g m^{-2}, the relationship with surface temperature is more linear compared to the mean LWP values, but the results are not as statically significant as the values obtained from the analysis poleward of 70°N (NSA, r = 0.359, p = 0.052; poleward of 70°N, r = 0.776, p = 0.000). The CMBE-NSA observations have a frequency of LWP clouds in excess of the threshold of 30-40%, and the CNRM-CM5 model has a very similar frequency and a similar surface temperature. MIROC5 has a frequency of opaque clouds greater than the CMBE-NSA data and surface temperatures greater than the CMBE-NSA data. CAM5 and HadGEM2-A have a frequency of LWP greater than 20 g m^{-2} less than 15%, but CAM5 has surface temperatures lower the CMBE-NSA observations whereas HadGEM2-A has surface temperatures greater than the observations.

When examining the frequency of clouds with IWPs greater than 25 g m^{-2}, there is not a statistically significant relationship between this frequency and surface temperature, but some insights about the models are discovered. For example, the HadGEM2-A model has a relatively high frequency of IWPs greater than 25 g m^{-2}, which may help explain why the surface temperature is relatively high even though the LWP threshold frequency is low. In fact, when examining the frequency of clouds with LWPs greater than 20 g m^{-2} or IWPs greater than 25 g m^{-2}, the linear regression line has a higher r-value (r = 0.435) than either the LWP or IWP threshold regressions.

Because of the contemporaneous nature of LWP and surface temperature data from the observations and models at NSA, one can stratify temperature biases according to the signs of the LWP bias in order to obtain a better sense of contribution of LWP biases to surface temperature biases at the NSA. Figure 9

examines the surface temperature bias during 4 distinct periods of Artic liquid cloud production compared to observations: (1) periods in which the models produce clouds with LWPs greater than 20 g m^{-2} and the observations have clouds with LWPs less than 20 g m^{-2}, (2) periods in which the models and observations both have LWPs greater than 20 g m^{-2}, (3) periods in which the models and observations both have LWPs less than 20 g m^{-2}, and (4) periods in which the models have LWPs less than 20 g m^{-2} and the observations have LWPs greater than 20 g m^{-2} (Figure 9). Considering the results in Figure 8, analyzing the IWPs would have been beneficial, but IWP data at the NSA do not exist for this time period. Only day two hindcast results are shown in Figure 9 for clarity.

MIROC5 has a mean surface temperature that is higher than the CMBE-NSA observations (e.g., Figure 6), and these higher values occur during periods in which MIROC5 has LWPs greater than 20 g m^{-2} and the observations have LWPs less than 20 g m^{-2}. In addition, higher surface temperatures compared to CMBE-NSA data occur when both MIROC5 and CMBE-NSA data have LWPs less than 20 g m^{-2}. To analyze the relative contribution of each of these errors to the time-mean bias in surface temperature, we compute a weighted bias by multiplying the percentage of time that model occurs in the regime by the temperature bias in that regime. For MIROC5, the weighted bias during the period in which the model has LWPs greater than 20 g m^{-2} and the observations have LWPs less than 20 g m^{-2} is greater than the regime when both the MIROC5 and the CMBE-NSA data have LWPs less than 20 g m^{-2}.

The surface temperatures in CAM4, CAM5 and IPSL-CM5A-LR are all lower than the CMBE-NSA observations (Figure 6). The largest absolute values of the weighted bias for the models occur when the observations have LWPs greater than 20 g m^{-2} and the models have LWPs less than the threshold, and surface temperature biases are statistically different than zero for all of these models in this period (Figure 9). The IPSL-CM5A-LR model also has a statistically significant difference when the model and CMBE-NSA observations both have LWPs greater than 20 g m^{-2}, but the weighted difference is not as large. These results for CAM4, CAM5, IPSL-CM5A-LR, CNRM-CM5, and MIROC5 show that the surface temperature

biases largely relate to when the models fail to produce the liquid cloud state when it was observed.

The HadGEM2-A model is an exception at NSA. When compared to the CMBE-NSA data, the HadGEM2-A model has a positive surface temperature bias. The largest weighted bias in HadGEM2-A occurs during periods in which both HadGEM2-A and the observations have LWPs less than 20 g m^{-2}. As shown in the IWP and IWP frequency scatter plots, the HadGEM2-A has relatively larger values and more frequent periods of large IWPs. During the period in which the HadGEM2-A and CMBE-NSA observations had LWPs less than 20 g m^{-2}, HadGEM2-A IWPs were 43.4 g m^{-2} , while every other model had IWPs less than 23 g m^{-2}. Still considering how the warm bias in surface at NSA is not representative of the biases over larger areas in HadGEM2-A (Figure 2), it is not clear how much emphasis should be placed on the results for this model at NSA.

5.3. *Free-Running Model (AMIP) Biases in Arctic LTS*

5.3.1. *Biases in LTS, Potential Temperature at 850hPa, and Near Surface Temperature*

How do these hindcast biases relate to biases of the same model run in AMIP mode? Xie et al. (2012), Williams et al. (2013), and Ma et al. (2014) determined that many fast physics AMIP modeling errors occur in hindcast simulations, but Arctic LTS was not specifically examined.

For the AMIP analysis, biases in each month are examined because data for all months are available, and there are some similarities and differences between the biases in AMIP (Figure 10) and hindcast (Figure 1) mode. Similar to the hindcast runs, AMIP runs of CAM4, CAM5, and HadGEM2-A have larger LTS values compared to the ERA-I reanalysis during the winter months. The winter month LTS biases in CAM4, CAM5, and HadGEM2-A are largely due to the surface temperature being lower than ERA-I. As with the hindcast modeling runs, difficulties in representing Arctic LTS during the winter is related to difficulties in representing the surface temperature. CNRM-CM5 has similar LTS and surface temperature compared to the ERA-I reanalysis during the winter months, which is similar to the hindcast modeling results.

653 However, there are differences between the AMIP results and hindcast
654 results including an enhanced LTS error during the summer months in CAM4, CAM5,
655 HadGEM2-A, and CNRM-CM5 in the AMIP simulations; and the IPSL-CM5-LR and
656 MIROC5 AMIP simulations having similar LTS and surface temperature values
657 compared to ERA-I, whereas large biases occurred in the hindcast runs. In CAM4,
658 CAM5, and HadGEM2-A, the summer LTS bias is a combination of the biases at 850
659 hPa and the surface. In these models, the surface is slightly colder than the ERA-I
660 reanalysis and the temperature at 850 hPa is slightly larger. In the CNRM-CM5
661 model, the summer LTS bias is largely due to the potential temperature at 850 hPa.
662 In 4 of the 6 models (CAM4, CAM5, HadGEM2-A, and CNRM-CM5), the
663 winter/polar night LTS bias was similar in the AMIP runs compared to the hindcast
664 runs. MIROC5 and IPSL-CM5A-LR have dissimilar biases between AMIP and hindcast
665 during the winter period. The main conclusion of Xie et al. (2012), which states fast
666 physics (e.g., clouds, radiation) hindcast errors is well related to those errors
667 manifest in AMIP simulations, seems only partially confirmed when examining the
668 LTS errors in the Arctic region, but a larger sample size of different GCMs would be
669 beneficial.

670 **5.3.1.1.** ***Comparison with LW Radiation***

671 To further compare the AMIP runs to the hindcast runs, we focus on polar
672 night biases defined by daily data, and only examine output over sea ice. Surface
673 temperature biases are compared to LW↓, clear sky LW↓, and CRE_{LW} (Figure 11).
674 The regression between surface temperature and LW↓ is 0.573 during the polar
675 night, and the slope is 2.32 W m^{-2} K^{-1}, which indicates that more processes than LW
676 radiative equilibrium are controlling the surface temperature. The clear sky LW↓
677 and CRE_{LW} do not significantly relate to surface temperature. Interestingly, the
678 CRE_{LW} r value is more statistically significant than the clear sky LW↓ relationship,
679 which is dissimilar to the hindcast simulations. Although the signs of the regressions
680 between AMIP surface temperatures and cloud properties during the polar night
681 period match those of the hindcasts, they are not statistically significant (not
682 shown).

683 The differences between the AMIP and hindcast relationships are not
684 surprising because of the multiple feedbacks that occur in AMIP simulations. For
685 example, Song and Mapes (2012) suggested that day 30 and greater hindcast Arctic
686 errors in the Climate Forecasting Model are related to errors in the thermal wind,
687 which relates to the advection of heat and moisture into the Arctic.

688 **6. Conclusions**

689 This study examined Global Climate Models (GCMs) run in hindcast and free
690 running mode to analyze errors in the Artic lower tropospheric stability. Because
691 Arctic lower tropospheric stability may be related Arctic climate change (Bintanja et
692 al. 2011; Bintanja et al. 2012; Boe et al. 2009), it is important to have a better
693 understanding of lower tropospheric stability errors and hypothesize why these
694 errors exist. The results related to the questions posed in the introduction are
695 summarized below.

696 - *Do GCMs have biases in Arctic LTS during all months/seasons?* We found that
697 the Arctic lower tropospheric stability bias predominantly occurs during the
698 polar night over sea ice regions in the hindcast simulations. AMIP
699 simulations do not show as clear of a seasonality in lower tropospheric
700 stability biases as the hindcast simulations, but similar biases occur during
701 the winter months.
702 - *What atmospheric level or levels may be causing an Arctic bias in LTS?* The
703 polar night Arctic lower tropospheric stability bias is related to biases in
704 Arctic surface temperature. This is displayed in the hindcast and AMIP
705 modeling runs. AMIP summer biases relate to biases at the surface as well as
706 temperature biases above the inversion layer.
707 - *What are possible causes of the Arctic LTS bias?* Downwelling longwave
708 radiation significantly relates to the models' spread in surface temperature
709 during the polar night in the hindcast simulations. Clear sky longwave
710 radiation and the cloud radiative effect both relate to the spread in surface
711 temperature in the hindcast models, and the production of cloud liquid water
712 is shown to significantly relate to surface temperatures over the Arctic

713 region. Ice water in Arctic clouds is important in the HadGEM2-A model
714 when analyzing hindcast simulations at the North Slope of Alaska.
715 The contribution of longwave radiation on surface temperature, hence lower
716 tropospheric stability, in multiple models is very evident in hindcast simulations,
717 but other variables must be affecting the model spread in AMIP/free running
718 simulations. In fact, Song and Mapes (2012) suggested circulation errors contribute
719 to Arctic surface temperature errors in 30 day hindcasts of a fully coupled
720 atmosphere-ocean-land model, which is well beyond the up to 5 day hindcasts
721 examined here. Though beyond the scope of the T-AMIP modeling runs, it would be
722 beneficial to determine if the same models that have Arctic temperature errors in
723 the 5 day hindcast have similar errors at time periods in which the dynamics have a
724 greater chance to feed back onto the system.
725 The analysis shows strong co-variability between clear sky longwave
726 radiation, the cloud radiative effect, largely due to differences in liquid clouds, and
727 surface temperature during the polar night periods. Models that produced larger
728 values clear sky LW also produced larger CRE_{LW} values and a higher frequency of
729 clouds with LWPs greater than 20 g m^{-2}. There is interplay between the clear sky
730 and clouds that affect the Arctic polar night surface temperature in these T-AMIP
731 models. The connection between clear sky temperature, moisture, and clouds
732 greatly affects the Arctic surface radiative budget (Francis et al. 2005; Stramler et al.
733 2011; Morrison et al. 2012), but it is not clear if clear sky periods or cloudy periods
734 drives the relationship.
735 It is well-known that models have difficulties in reproducing Arctic clouds
736 (Tjernstrom et al. 2008; Klein et al. 2009; Morrison et al. 2009), and we show that
737 the frequency of opaqueness is an important variable when determining the clouds
738 effect on the radiation budget in models. As shown in the observational studies of
739 Stramler et al. (2011) and Morrison et al. (2012), the distribution of variables that
740 affect surface longwave radiation in models is important for the overall Arctic
741 surface radiative budget. Pithan et al. (2013) suggested the misrepresentations of
742 mixed-phase clouds relates to biases in Arctic LTS in the CMIP5 archive. We add that
743 in addition to cloud biases, clear sky longwave biases are also important when

744 examining lower tropospheric stability errors. Improved simulation of the processes
745 controlling downward longwave radiation in polar night will lead to an improved
746 simulation of Arctic climate. Improved downward longwave radiation will require
747 an attention to mixed-phase clouds, frequency of liquid that creates opaque Arctic
748 clouds, and processes controlling clear sky longwave radiation.
749 **ACKNOWLEDGEMENTS:**
750 The contribution of N.P. Barton, S.A. Klein, and J.S. Boyle to this work was performed
751 under the auspices of the U.S. Department of Energy by Lawrence Livermore
752 National Laboratory under contract DE-AC52-07NA27344. Support for N.P. Barton,
753 S.A. Klein, and J.S. Boyle was provided by the Regional and Global Climate and Earth
754 System Modeling Programs of the Office of Science at the U. S. Department of
755 Energy. We acknowledge the Working Group on Numerical Experiments (WGNE)
756 and the Working Group on Coupled Modeling (WGCM) who are responsible for
757 Transpose-AMIP II, and we thank the climate modeling groups (listed in Table 1 of
758 this paper) for producing and making available their model output. For Transpose-
759 AMIP II the U.S. Department of Energy's Program for Climate Model Diagnosis and
760 Intercomparison (PCMDI) provides coordinating support and led development of
761 software infrastructure in partnership with the Global Organization for Earth
762 System Science Portals.
763
764
765

766 **REFERENCES:**

767 Barton, N. P., and D. E. Veron, 2013: Response of clouds and surface energy fluxes to
768 changes in sea-ice cover over the Laptev Sea (Arctic Ocean). *Climate Research*, **54,**
769 69-84.

770 Barton, N. P., S. A. Klein, J. S. Boyle, and Y. Y. Y. Zhang, 2012: Arctic synoptic regimes:
771 Comparing domain-wide Arctic cloud observations with CAM4 and CAM5 during
772 similar dynamics. *Journal of Geophysical Research-Atmospheres*, **117,** 19.

773 Bintanja, R., R. G. Graversen, and W. Hazeleger, 2011: Arctic winter warming
774 amplified by the thermal inversion and consequent low infrared cooling to space.
775 *Nature Geoscience*, **4,** 758-761.

776 Bintanja, R., E. C. van der Linden, and W. Hazeleger, 2012: Boundary layer stability
777 and Arctic climate change: a feedback study using EC-Earth. *Climate Dynamics*,
778 **39,** 2659-2673.

779 Boe, J., A. Hall, and X. Qu, 2009: Current GCMs' Unrealistic Negative Feedback in the
780 Arctic. *Journal of Climate*, **22,** 4682-4695.

781 Chapman, W. L., and J. E. Walsh, 2007: Simulations of Arctic temperature and
782 pressure by global coupled models. *Journal of Climate*, **20,** 609-632.

783 Chen, Y. H., F. Aires, J. A. Francis, and J. R. Miller, 2006: Observed relationships
784 between arctic longwave cloud forcing and cloud parameters using a neural
785 network. *Journal of Climate*, **19,** 4087-4104.

786 Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and
787 performance of the data assimilation system. *Quarterly Journal of the Royal*
788 *Meteorological Society*, **137,** 553-597.

789 Derbyshire, S. H., 1999: Stable boundary-layer modelling: Established approaches
790 and beyond. *Boundary-Layer Meteorology*, **90,** 423-446.

791 Dethloff, K., C. Abegg, A. Rinke, I. Hebestadt, and V. F. Romanov, 2001: Sensitivity of
792 Arctic climate simulations to different boundary-layer parameterizations in a
793 regional climate model. *Tellus Series a-Dynamic Meteorology and Oceanography*,
794 **53,** 1-26.

795 Ebert, E. E., and J. A. Curry, 1992: A parameterization of ice-cloud optical-properties
796 for climate models. *Journal of Geophysical Research-Atmospheres*, **97,** 3831-3836.

797 Francis, J. A., E. Hunter, J. R. Key, and X. J. Wang, 2005: Clues to variability in Arctic
798 minimum sea ice extent. *Geophysical Research Letters*, **32,** 4.

799 Garrett, T. J., and C. F. Zhao, 2006: Increased Arctic cloud longwave emissivity
800 associated with pollution from mid-latitudes. *Nature*, **440,** 787-789.

801 Kay, J. E., and A. Gettelman, 2009: Cloud influence on and response to seasonal
802 Arctic sea ice loss. *Journal of Geophysical Research-Atmospheres*, **114**.

803 Kay, J. E., K. Raeder, A. Gettelman, and J. Anderson, 2011: The Boundary Layer
804 Response to Recent Arctic Sea Ice Loss and Implications for High-Latitude
805 Climate Feedbacks. *Journal of Climate*, **24,** 428-447.

806 Klein, S. A., and D. L. Hartmann, 1993: The Seasonal Cycle of Low Stratiform Clouds.
807 *Journal of Climate*, **6,** 1587-1606.

808 Klein, S. A., and Coauthors, 2009: Intercomparison of model simulations of mixed-
809 phase clouds observed during the ARM Mixed-Phase Arctic Cloud Experiment. I:
810 Single-layer cloud. *Quarterly Journal of the Royal Meteorological Society*, **135,**
811 979-1002.

812 Liu, X., and Coauthors, 2011: Testing cloud microphysics parameterizations in NCAR
813 CAM5 with ISDAC and M-PACE observations. *J. Geophys. Res.*, **116,** D00T11.

814 Liu, Y. H., J. R. Key, A. Schweiger, and J. Francis, 2006: Characteristics of satellite-
815 derived clear-sky atmospheric temperature inversion strength in the Arctic,
816 1980-96. *Journal of Climate*, **19,** 4902-4913.

817 Ma, H. Y., S. Xie, J. S. Boyle, S. A. Klein, and Y. Zhang, 2013: Metrics and Diagnostics
818 for Precipitation-Related Processes in Climate Model Short-Range Hindcasts.
819 *Journal of Climate*, **26,** 1516-1534.

820 Ma, H. Y., and Coauthors, 2014: On the Correspondence between Mean Forecast
821 Errors and Climate Errors in CMIP5 Models. *Journal of Climate*, **27,** 1781-1798.

822 Medeiros, B., C. Deser, R. A. Tomas, and J. E. Kay, 2011: Arctic Inversion Strength in
823 Climate Models. *Journal of Climate*, **24,** 4733-4740.

824 Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997:
825 Radiative transfer for inhomogeneous atmospheres: RRTM, a validated
826 correlated-k model for the longwave. *Journal of Geophysical Research-*
827 *Atmospheres*, **102,** 16663-16682.

828 Morrison, H., G. de Boer, G. Feingold, J. Harrington, M. D. Shupe, and K. Sulia, 2012:
829 Resilience of persistent Arctic mixed-phase clouds. *Nature Geoscience*, **5,** 11-17.

830 Morrison, H., and Coauthors, 2009: Intercomparison of model simulations of mixed-
831 phase clouds observed during the ARM Mixed-Phase Arctic Cloud Experiment. II:
832 Multilayer cloud. *Quarterly Journal of the Royal Meteorological Society*, **135,** 1003-
833 1019.

834 Oreopoulos, L., and Coauthors, 2012: The Continual Intercomparison of Radiation
835 Codes: Results from Phase I. *Journal of Geophysical Research-Atmospheres*, **117,**
836 19.

837 Pavelsky, T. M., J. Boe, A. Hall, and E. J. Fetzer, 2011: Atmospheric inversion strength
838 over polar oceans in winter regulated by sea ice. *Climate Dynamics*, **36,** 945-955.

839 Phillips, T. J., and Coauthors, 2004: Evaluating parameterizations in general
840 circulation models - Climate simulation meets weather prediction. *Bulletin of the*
841 *American Meteorological Society*, **85,** 1903-1915.

842 Pithan, F., B. Medeiros, and T. Mauritsen, 2013: Mixed-phase clouds cause climate
843 model biases in Arctic wintertime temperature inversions. *Climate Dynamics*, 1-
844 15.

845 Ramanathan, V., R. D. Cess, E. F. Harrison, P. Minnis, B. R. Barkstrom, E. Ahmad, and
846 D. Hartmann, 1989: Cloud-radiative forcing and climate- results from the Earth
847 Radiation Budget Experiment. *Science*, **243,** 57-63.

848 Sandu, I., A. Beljaars, P. Bechtold, T. Mauritsen, and G. Balsamo, 2013: Why is it so
849 difficult to represent stably stratified conditions in numerical weather prediction
850 (NWP) models? *Journal of Advances in Modeling Earth Systems*, n/a-n/a.

851 Serreze, M. C., J. D. Kahl, and R. C. Schnell, 1992: Low-level Temperature Inversions
852 of the Eurasian Arctic and Comparisons with Soviet Drifting Station Data. *Journal*
853 *of Climate*, **5,** 615-629.

854 Shupe, M. D., and J. M. Intrieri, 2004: Cloud radiative forcing of the Arctic surface:
855 The influence of cloud properties, surface albedo, and solar zenith angle. *Journal*
856 *of Climate*, **17,** 616-628.

857 Shupe, M. D., V. P. Walden, E. Eloranta, T. Uttal, J. R. Campbell, S. M. Starkweather,
858 and M. Shiobara, 2011: Clouds at Arctic Atmospheric Observatories. Part I:
859 Occurrence and Macrophysical Properties. *Journal of Applied Meteorology and*
860 *Climatology*, **50,** 626-644.

861 Song, S. W., and B. Mapes, 2012: Interpretations of systematic errors in the NCEP
862 Climate Forecast System at lead times of 2, 4, 8, ..., 256 days. *Journal of Advances*
863 *in Modeling Earth Systems*, **4,** 11.

864 Steeneveld, G. J., B. J. H. Van de Wiel, and A. A. M. Holtslag, 2006: Modelling the arctic
865 stable boundary layer and its coupling to the surface. *Boundary-Layer*
866 *Meteorology*, **118,** 357-378.

867 Stramler, K., A. D. Del Genio, and W. B. Rossow, 2011: Synoptically Driven Arctic
868 Winter States. *Journal of Climate*, **24,** 1747-1762.

869 Svensson, G., and J. Karlsson, 2011: On the Arctic Wintertime Climate in Global
870 Climate Models. *Journal of Climate*, **24,** 5757-5771.

871 Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: AN OVERVIEW OF CMIP5 AND THE
872 EXPERIMENT DESIGN. *Bulletin of the American Meteorological Society*, **93,** 485-
873 498.

874 Tjernstrom, M., and R. G. Graversen, 2009: The vertical structure of the lower Arctic
875 troposphere analysed from observations and the ERA-40 reanalysis. *Quarterly*
876 *Journal of the Royal Meteorological Society*, **135,** 431-443.

877 Tjernstrom, M., J. Sedlar, and M. D. Shupe, 2008: How well do regional climate
878 models reproduce radiation and clouds in the Arctic? An evaluation of ARCMIP
879 simulations. *Journal of Applied Meteorology and Climatology*, **47,** 2405-2422.

880 Waliser, D. E., and Coauthors, 2012: THE "YEAR" OF TROPICAL CONVECTION (MAY
881 2008-APRIL 2010) Climate Variability and Weather Highlights. *Bulletin of the*
882 *American Meteorological Society*, **93,** 1189-1218.

883 Westwater, E. R., Y. Han, M. D. Shupe, and S. Y. Matrosov, 2001: Analysis of
884 integrated cloud liquid and precipitable water vapor retrievals from microwave
885 radiometers during the Surface Heat Budget of the Arctic Ocean project. *Journal*
886 *of Geophysical Research-Atmospheres*, **106,** 32019-32030.

887 Williams, K. D., and Coauthors, 2013: The Transpose-AMIP II Experiment and Its
888 Application to the Understanding of Southern Ocean Cloud Biases in Climate
889 Models. *Journal of Climate*, **26,** 3258-3274.

890 Xie, S., J. Boyle, S. A. Klein, X. Liu, and S. Ghan, 2008: Simulations of Arctic mixed-
891 phase clouds in forecasts with CAM3 and AM2 for M-PACE. *Journal of Geophysical*
892 *Research-Atmospheres*, **113**.

893 Xie, S., and Coauthors, 2010: CLOUDS AND MORE: ARM Climate Modeling Best
894 Estimate Data. *Bulletin of the American Meteorological Society*, **91,** 13-20.

895 Xie, S. C., H. Y. Ma, J. S. Boyle, S. A. Klein, and Y. Y. Zhang, 2012: On the
896 Correspondence between Short- and Long-Time-Scale Systematic Errors in
897 CAM4/CAM5 for the Year of Tropical Convection. *Journal of Climate*, **25,** 7937-
898 7955.

899 Zhang, Y. H., D. J. Seidel, J. C. Golaz, C. Deser, and R. A. Tomas, 2011: Climatological
900 Characteristics of Arctic and Antarctic Surface-Based Inversions. *Journal of*
901 *Climate*, **24,** 5167-5186.

902

TABLES:

Table 1: Model reference for each GCM used in this study.

Model	Reference
CAM4	*Gent et al.,* 2011
CAM5	*Neale et al.,* 2010
HadGEM2-A	*Martin et al.,* 2011
IPSL-CM5A-LR	http://icmc.ipsl.fr
CNRM-CM5	*Voldoire et al.,* 2013
MIROC5	*Watanabe et al.,* 2010

Table 2: Biases, Mean Absolute Errors (MAE), and Root Mean Squared Errors (RMSE) for Lower Tropospheric Stability (LTS), potential temperature at 850 hPa (θ_{850}), and near surface temperature ($\theta_{surface}$) between the ECMWF-Y analysis and CMBE-NSA data for all time periods in the analysis, polar night periods, and polar day periods. LTS is defined by Eq. 1, and polar night and polar day are defined in section 4.1. All units are in degree K.

	Bias			MAE			RMSE		
	All	Polar Night	Polar Day	All	Polar Night	Polar Day	All	Polar Night	Polar Day
LTS	-0.36	0.40	-1.00	2.02	2.63	1.51	2.82	3.55	1.97
$\boldsymbol{\theta_{850}}$	-0.11	-0.13	-0.09	0.87	1.02	0.75	1.52	1.85	1.18
$\boldsymbol{\theta_{surface}}$	0.11	-0.46	0.68	1.56	1.72	1.40	2.07	2.35	1.75

FIGURES:

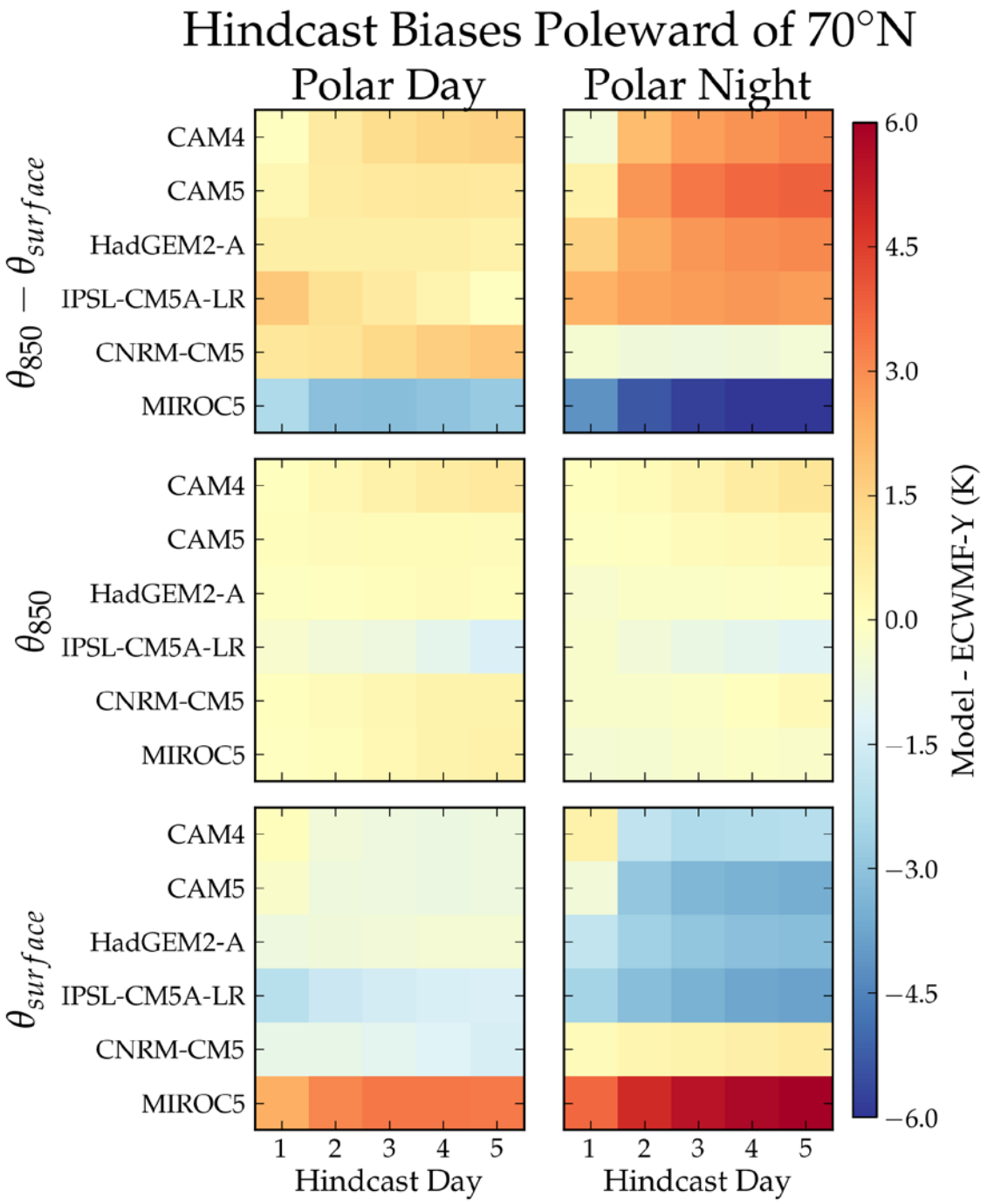

Figure 1: Hindcast model biases of Arctic region (top) lower tropospheric stability, (middle) potential temperature at 850 hPa, and (bottom) surface temperature averaged for periods of (left) Polar Day and (right) Polar Night. The Arctic is defined as poleward of 70°N. The biases are calculated from the ECMWF-Y analysis. The y-axis displays the bias for each day since start of the hindcast, and the x-axis displays the bias for each model.

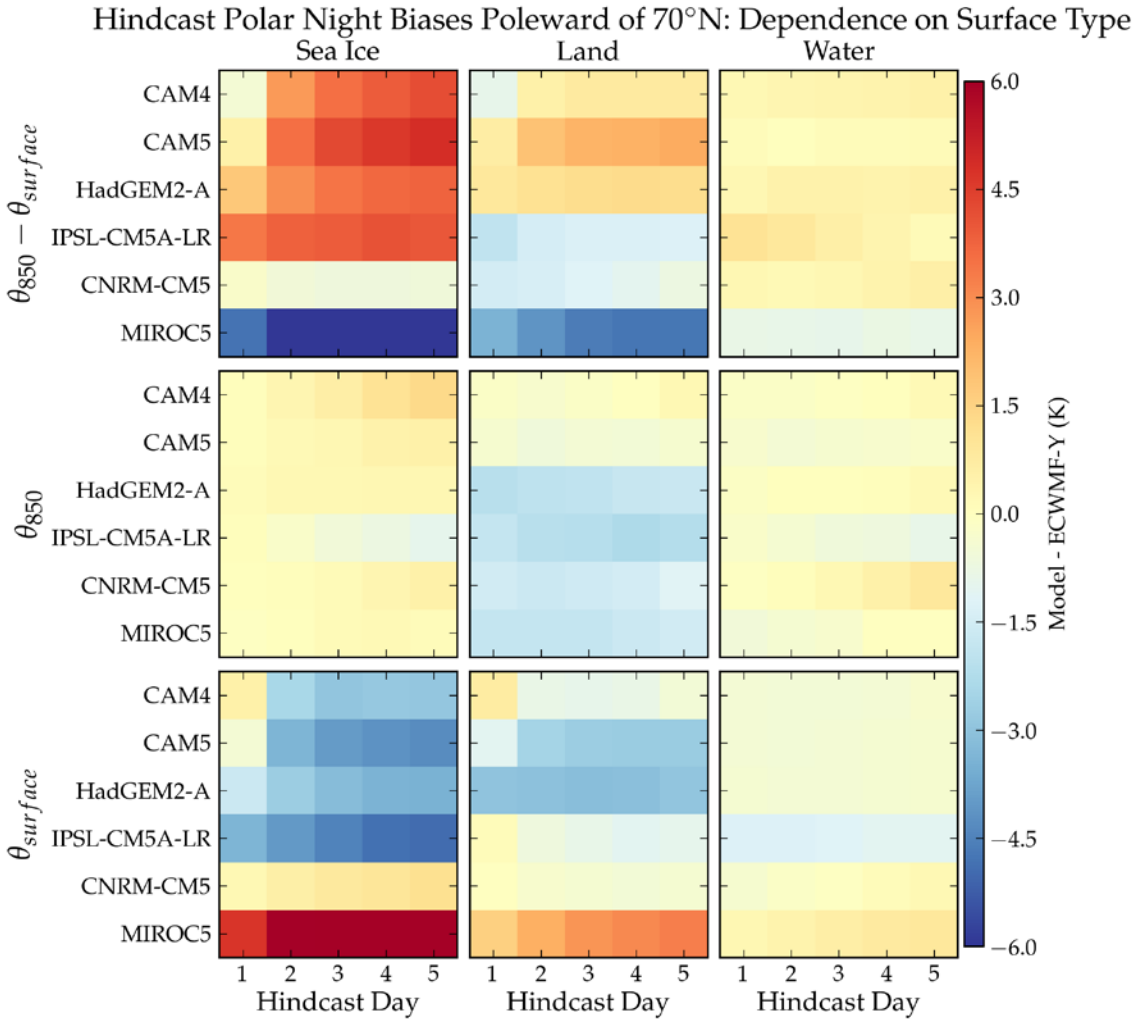

Figure 2: Hindcast model biases of Arctic region (top) lower tropospheric stability, (middle) potential temperature at 850 hPa, and (bottom), surface temperature averaged for Polar Night periods averaged over (left) sea ice, (middle) land, and (right) water. The biases are calculated from the ECMWF-Y analysis. The y-axis displays the bias for different hindcast days, and the x-axis displays the bias for each model. Sea ice regions are defined by concentrations greater than 15%.

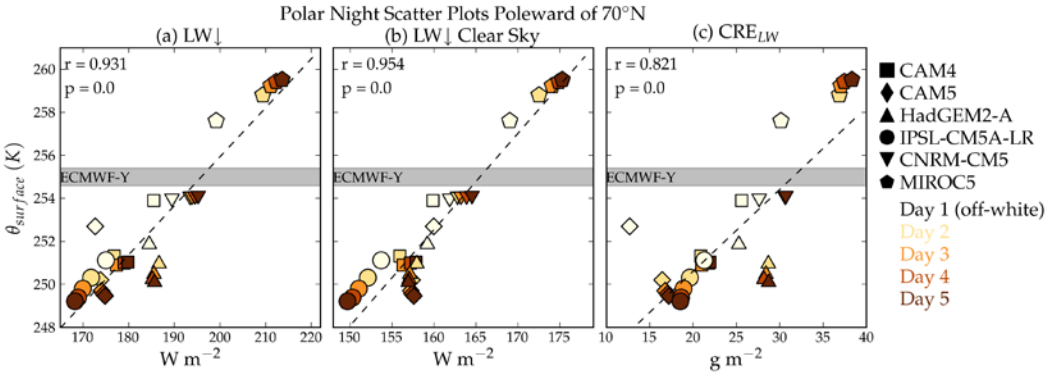

Figure 3: Scatter plots between surface temperature and (a) downwelling longwave radiation at the surface (LW↓), (b) clear sky downwelling longwave radiation at the surface (LW↓ Clear Sky), and (c) the surface longwave cloud radiative effect (CRE_{LW}) for the hindcast model runs. All values are averaged poleward of 70°N over sea ice regions defined by 15% ice concentration. Each model is represented with a different symbol: (square) CAM4, (diamond) CAM5, (point-up triangle) HadGEM2-A, (circle) CNRM-CM5, (point-down triangle) CNRM-CM5, and (pentagon) MIROC5. Different shadings of each symbol represent the hindcast day, with darker shading representing latter forecast days. The grey bar represents the 95% confidence interval of surface temperatures from the ECMWF-Y analysis assuming a Gaussian distribution. The dash line is the best-fit linear regression between the two variables. The r and p-values from the linear regression is placed in the top left corner of each graph.

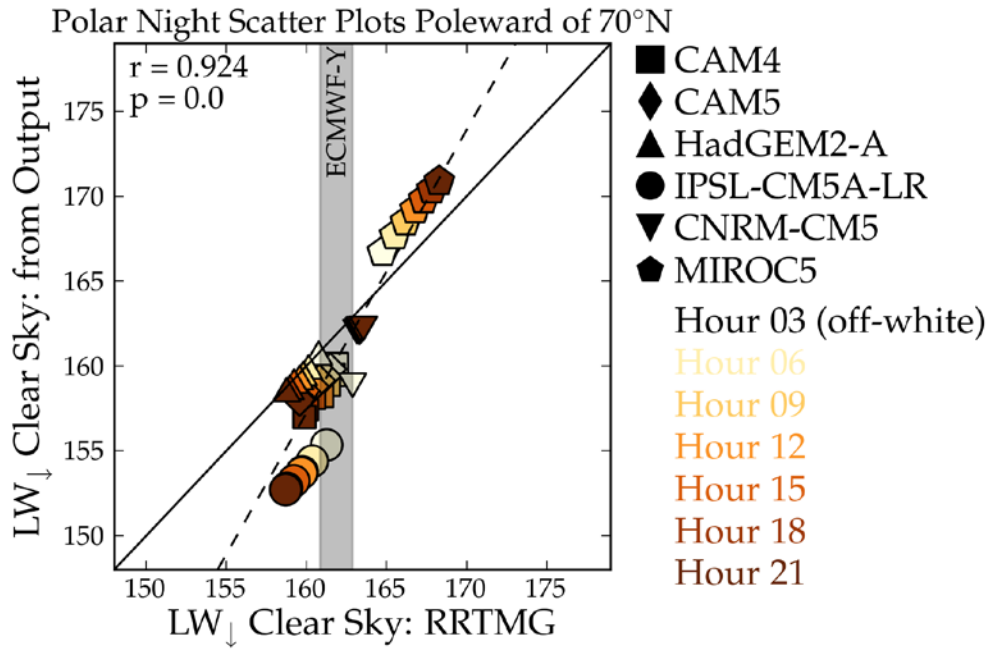

Figure 4: Same as Figure 3 except for clear sky longwave down output from each model is on the y-axis and clear sky longwave down output from RRTMG is on the x-axis, and hindcast hours 3 to 21 are displayed instead daily averages. The solid black

line is the one-to-one line. The black dot with error bars represents the clear sky longwave downwelling radiation from RRTMG using the ECMWF-Y analysis. The error bars represent the 95% confidence interval.

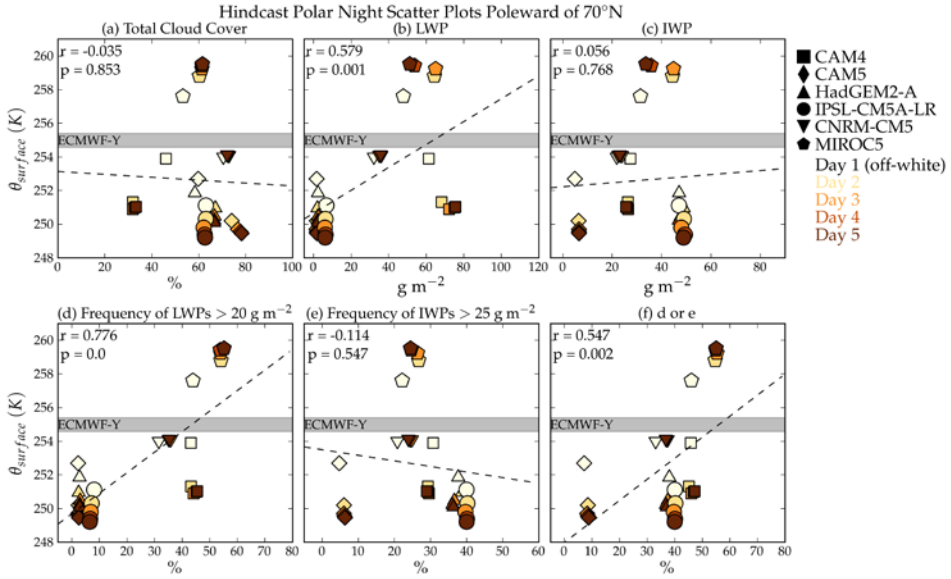

Figure 5: Same as Figure 3 except for (a) Total Cloud Cover, (b) all sky Liquid Water Path, (c) all sky Ice Water Path, (d) the frequency of the LWPs greater than 20 g m^{-2}, (e) the frequency of IWPs greater than 25 g m^{-2}, and (f) the frequency that LWPs were greater than 20 g m^{-2} or IWPs were greater than 25 g m^{-2}.

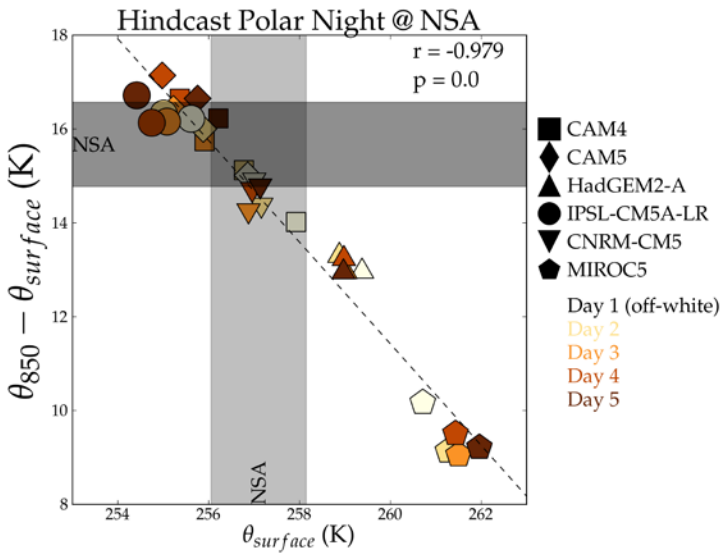

Figure 6: Similar to Figure 3, except the scatter between (y-axis) LTS and (x-axis) near surface temperature at the North Slope of Alaska (NSA).

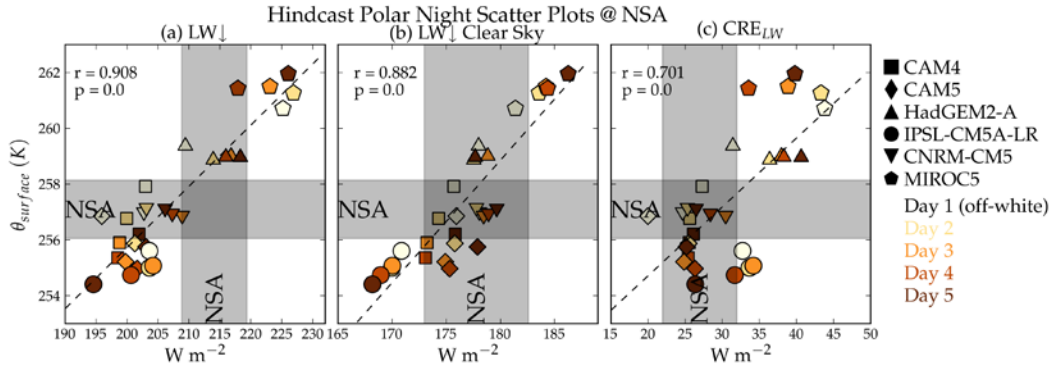

Figure 7: Same as Figure 3 except the analysis is performed at the North Slope of Alaska (NSA).

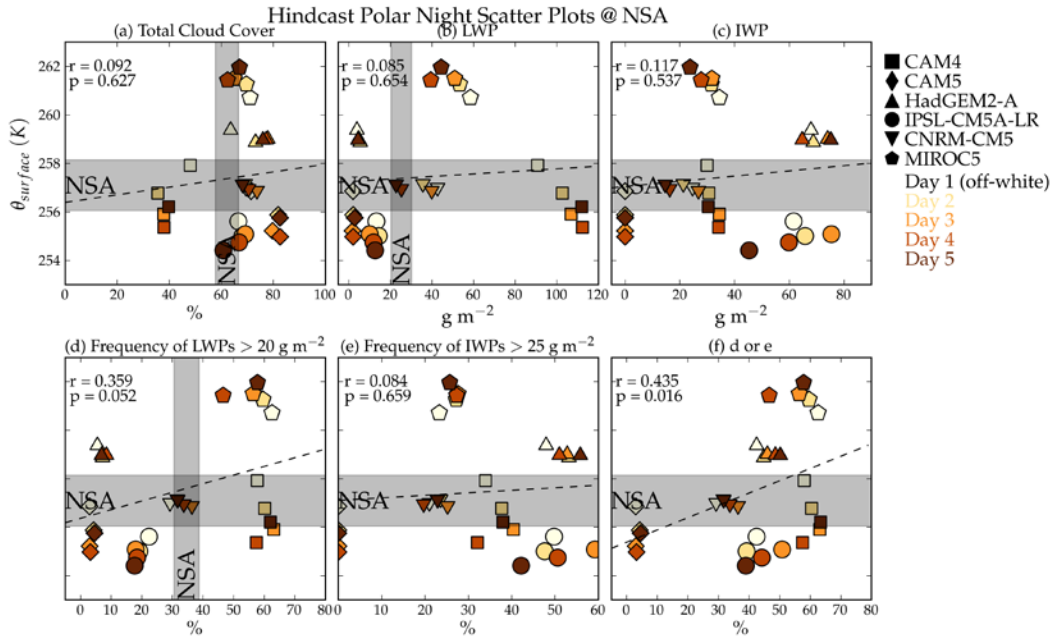

Figure 8: Same as Figure 5, except the analysis is performed at the North Slope of Alaska (NSA). There are not Ice Water Path data at the NSA for this time period. The confidence interval for the NSA frequency of LWPs greater than 20 g m^{-2} was computed using bootstrapping by removing a data point and recalculating the frequency 1,000 times.

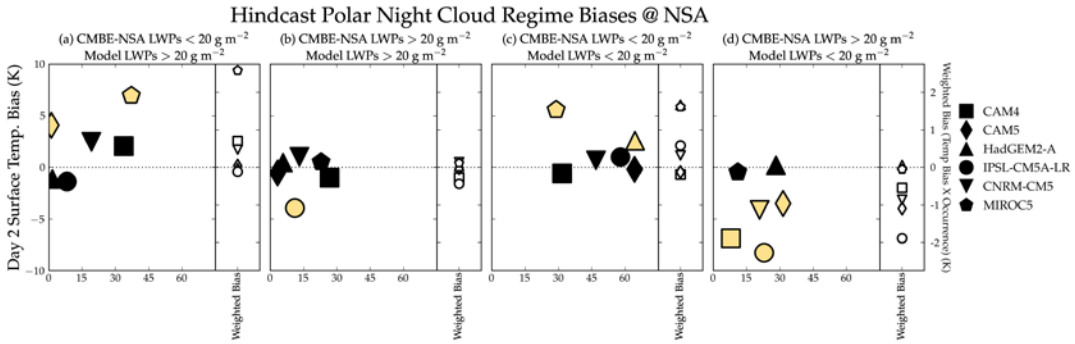

Figure 9: Day two temperature biases for four periods of modeled cloud production compared to observations at the NSA: (a) periods in which the models produce clouds with LWPs greater than 20 g m^{-2} and the observations have clouds with LWPs less than 20 g m^{-2}, (b) periods in which the models and observations have clouds with LWPs greater than 20 g m^{-2}, (c) periods in which the models and observations have clouds with LWPs less than 20 g m^{-2}, and (d) periods in which the models produced clouds with LWPs less than 20 g m^{-2} and the observations have clouds with LWPs greater than 20 g m^{-2}. The temperature bias is on the left y-axis

and the percent of occurrence that each model occurs in these regimes is the x-axis. The colored in symbols represent biases for the regime that have p values less than 0.05 defined by a t-tailed student t-test. The right y-axis is the weighted bias and is calculated by the bias multiplied by the percent of occurrence.

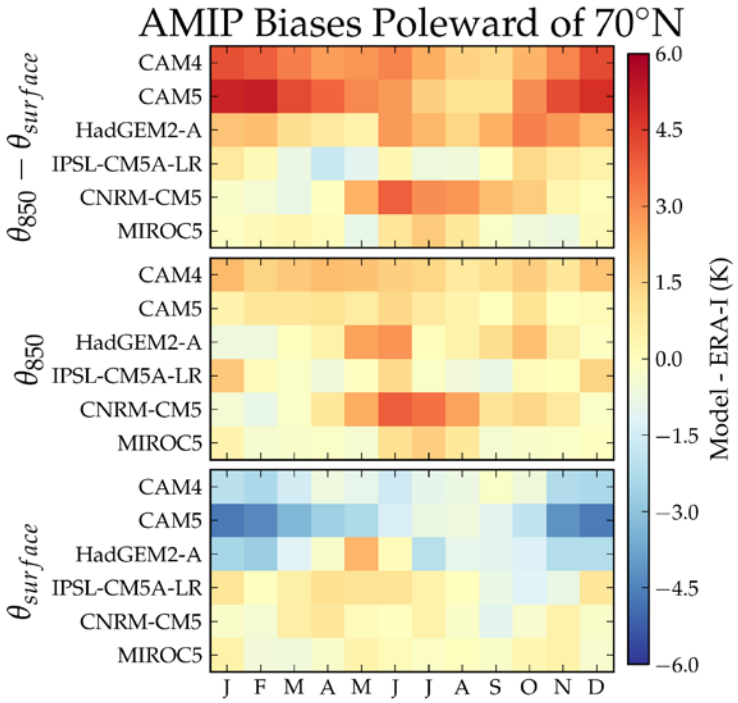

Figure 10: AMIP model biases for (top) lower tropospheric stability, (middle) potential temperature at 850 hPa, and (bottom) surface temperature. This biases are calculated from the ERA-I reanalysis. The vertical axis represents the individual model, and the horizontal axis is the month of the bias.

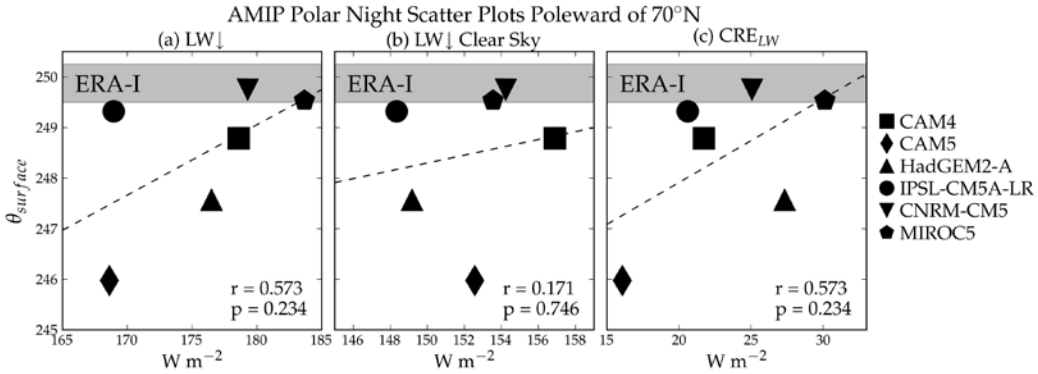

Figure 11: Same as Figure 3 except the analysis is performed on AMIP simulations from 1979 to 2008 using daily output poleward of 70°N and over sea ice regions. The gray shading represents the 95% confidence interval for the ERA-I reanalysis.